Getting more out of Area of Interest (AOI) analysis with SPLOT

Belopolsky, Artem V.

published in
ETRA 2020 Short Papers
2020

DOI (link to publisher)
10.1145/3379156.3391372

document version
Publisher's PDF, also known as Version of record

document license
Article 25fa Dutch Copyright Act

citation for published version (APA)
Belopolsky, A. V. (2020). Getting more out of Area of Interest (AOI) analysis with SPLOT. In S. N. Spencer (Ed.), *ETRA 2020 Short Papers: ACM Symposium on Eye Tracking Research and Applications* (pp. 1-4). [18] (Eye Tracking Research and Applications Symposium (ETRA)). Association for Computing Machinery. https://doi.org/10.1145/3379156.3391372

Getting more out of Area of Interest (AOI) analysis with SPLOT

Artem V. Belopolsky
Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
a.belopolskiy@vu.nl

ABSTRACT

To analyze eye-tracking data the viewed image is often divided into areas of interest (AOI). However, the temporal dynamics of eye movements towards the AOI is often lost either in favor of summary statistics (e.g., proportion of fixations or dwell time) or is significantly reduced by "binning" the data and computing the same summary statistic over each time bin. This paper introduces SPLOT: smoothed proportion of looks over time method for analyzing the eye movement dynamics across AOI. SPLOT comprises of a complete workflow, from visualization of the time-course to performing statistical analysis on it using cluster-based permutation testing. The possibilities of SPLOT are illustrated by applying it to an existing dataset of eye movements of radiologists diagnosing a chest X-ray.

CCS CONCEPTS

• **Applied computing** → Law, social and behavioral sciences.

KEYWORDS

Eye Tracking, Area Of Interest analysis, Eye movement dynamics, Cluster-based permutation, Proportion of looks over time

ACM Reference Format:
Artem V. Belopolsky. 2020. Getting more out of Area of Interest (AOI) analysis with SPLOT. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Short Papers), June 2–5, 2020, Stuttgart, Germany.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3379156.3391372

1 INTRODUCTION

Modern eye-trackers are capable of sampling eye position up to 2000 times a second. However, when studying complex oculomotor behaviors, the high temporal precision is often replaced by some form of summary statistics. For example, if one is interested in whether a human face appearing in a natural scene attracts attention, the location of a face is designated as an Area of Interest (AOI) and summary statistics, such as proportion of fixations on the face, the total dwell time, the time of first fixation, the number of returning saccadic eye movements, the number of transitions in and out of region, and many others are computed and aggregated over images and over participants [Holmqvist et al. 2011]. While this manipulation reduces the data complexity, the precious temporal information is also lost.

One common solution to preserve the dynamics of oculomotor behavior is to divide each trial in an arbitrary number of time intervals of equal size (e.g., to "bin" the data) and to compute the same summary statistic over each time bin [Andersson et al. 2011; De Groot et al. 2016]. While this solution is straightforward to implement, it suffers from a number of serious flaws. First, deciding on the number of bins is arbitrary and is usually done post-hoc, based on the number of available data points, and can significantly distort the underlying signal. Second, to avoid missing data, the temporal resolution is usually dramatically reduced and the bins are typically on the order of seconds, while the original signal is measured with a millisecond precision. Third, the measurements per bins are not independent from each other and should be corrected for multiple comparisons. This correction is quite conservative and makes it unappealing to have more bins (i.e., higher temporal precision), since the chances of observing significant results are decreasing with increasing number of bins.

In the present paper a new *Smoothed Proportion of Looks Over Time* method (SPLOT) is introduced, which allows to capture and analyze the eye movement dynamics for an AOI with high temporal precision. It is conceptually similar to the "*momentous proportion over time*" method [Holmqvist et al. 2011], but includes a complete workflow from data visualization to statistical analysis using permutation testing. It is inspired by a recently developed SMART method [van Leeuwen et al. 2019] which allows to reconstruct the time-course of behavioral data, sampled only once per trial, such as the relationship between accuracy and response time. In the following sections, the SPLOT method is described in detail and for illustration purposes, it is applied to a subset of eye movement dataset of radiologists diagnosing chest X-rays.

2 THE SPLOT METHOD

SPLOT consists of three main steps (Figure 1): 1) generating binary time-course; 2) temporal smoothing of trials and averaging them across participants; 3) cluster-based permutation testing.

2.1 Generating binary time-course

In order to reconstruct the eye movement time-course for a certain AOI, the eye movement sequence is first transformed into a sequence of fixations and their corresponding durations using any event-detection algorithm [Engbert and Mergenthaler 2006]. Then, fixations are transformed into "looks". The looks are coded as a binary variable, with '1' assigned to each time point that belongs to a fixation falling inside the AOI and '0' assigned to all other time points in the trial. Note, that saccades made within the AOI are also coded as '0's, since it is assumed that visual processing during a saccade is very limited [Matin 1974]. The total number of time points in a trial is determined by the sampling rate of the

ETRA '20 Short Papers, June 2–5, 2020, Stuttgart, Germany

ACM ISBN 978-1-4503-7134-6/20/06…$15.00
https://doi.org/10.1145/3379156.3391372

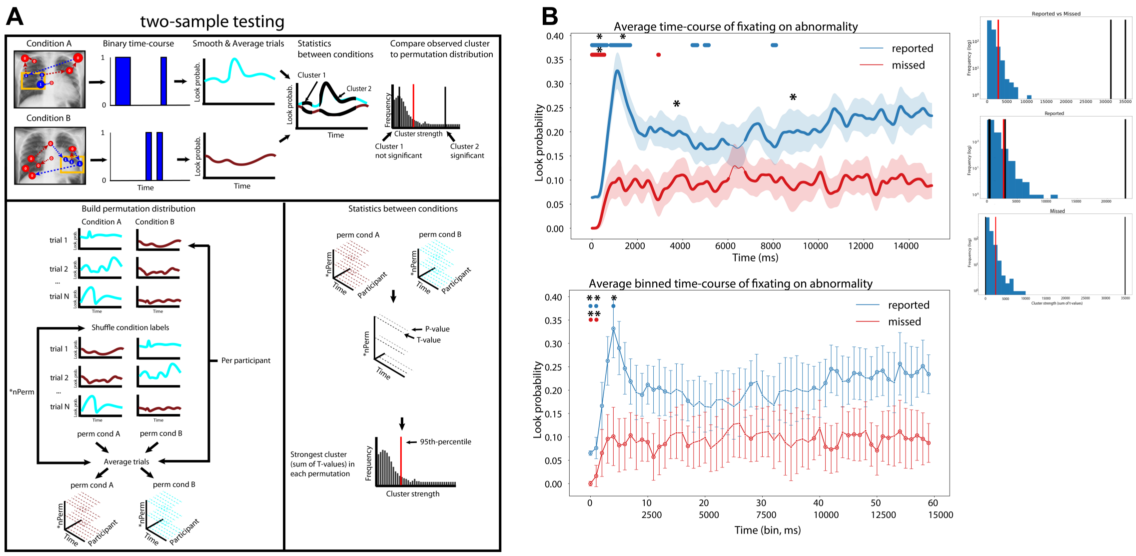

Figure 1: A. Illustration of the SPLOT method and cluster-based permutation for comparing the time-course between two conditions. B. Comparison of SPLOT and binning using real data. Top: SPLOT analysis with two- and one-sample permutation tests. Time points significant in reported vs. missed comparison are superimposed on the time-courses and significant clusters are indicated by a 'star'. Thick lines at the top indicate time points significant for one-sample test (against respective mean time-course) and significant clusters are indicated by a 'star'. Bottom: same data binned in 60 bins and analyzed using two- and one-sample t-tests with Bonferroni correction. All error bars are within-subject confidence intervals.

eye-tracker. This process is illustrated in Figure 1. For example, a participant in a study is viewing radiological images and each image is viewed for 15 seconds. The eye movements are sampled at 1000 Hz. Then each trial will contain 15000 data points. After transformation into looks, each fixation falling within the AOI will be represented by a square wave of "ones" that lasts as long as the fixation it represents, and all other data points will be zeros.

2.2 Temporal smoothing and averaging across trials

To reduce the noise present on individual trials and to convert a discreet signal into a continuous one, the square wave sequences are convolved with Gaussian kernel of a chosen size (Figure 1). The proposed kernel size is 100 ms, since this corresponds to minimal fixation duration size that reflects meaningful cognitive processing [Just and Carpenter 1980]. However, this parameter can be adjusted if needed. After each trial is smoothed, the trials can be averaged, producing an average time-course of proportion of looks at AOI for each participant. The advantage of SPLOT is that the time-courses can be averaged across different images, different AOIs and different participants. If there are similarities across looks, they should emerge in the average time-course.

2.3 Cluster-based permutation testing

Now that the oculomotor behavior on each trial has been converted into a continuous signal, the time-courses across different conditions can be statistically compared to each other to determine when the two signals are significantly different from each other. Furthermore, the time-courses for a single condition can be compared against a meaningful baseline, to determine when significant differences emerge. When performing a statistical test on each individual time point, clusters of significant points will emerge, since the time points are not independent of each other. Therefore, significance should be tested on the level of clusters, rather than on the level of individual data points. Cluster-based permutation testing provides a solution to this problem [van Leeuwen et al. 2019; Maris and Oostenveld 2007]. It involves building a distribution of the cluster-based test statistic under the null hypothesis and comparing the observed cluster-based statistic to it. Any cluster in the observed data, whose test statistic exceeds the 95th percentile (which corresponds to a p-value of 0.05) is considered significant. Since building a permutation distribution is different depending on whether two conditions are compared to each other or a single condition is compared to a baseline, these two procedures are described below.

2.3.1 Permutation testing between two conditions. In order to build a permutation distribution for testing between two conditions, each permutation iteration involves the following steps (Figure 1). First, the condition labels are randomly shuffled for each participant. Second, the time-courses for each condition are averaged for each participant. Third, a two-sample t-test is performed on each time point in the time-course. This can be a paired-sample t-test if conditions are manipulated within participants, but it can also be an independent-sample t-test if conditions are manipulated between participants. Any time point whose significance level exceeds p=.05 is considered significant. Any two or more adjacent significant time points are considered a cluster. Fourth, for each cluster, strength is determined by summing up the t-values belonging to a cluster. If a permutation contains no significant points, the largest t-value is used. The cluster strengths are added to the permutation distribution. These steps are repeated for each permutation iteration.

2.3.2 Permutation testing against a baseline. Sometimes it can be meaningful to compare a time-course against a baseline, such as an average time-course.

The first step in building a permutation distribution is to generate a baseline condition. The average baseline is generated by replacing the original time-course on every trial with a vector containing its mean (e.g., a flat line at the level of the mean). After that, the original and baseline conditions are compared using the permutation testing between two conditions procedure, which is described above.

3 APPLICATION TO THE EXISTING DATA

3.1 Methods

To illustrate the SPLOT method we apply it to a subset of the existing dataset of eye movement recordings of radiologists available in our lab. 25 radiologists from the VU University Medical Center, were invited to participate in an experiment, in which they were asked to diagnose a chest X-ray, while their eye movements were recorded. Eye movements were recorded with a desk-mounted remote EyeLink1000 system with 1000 Hz temporal and 0.01° spatial resolution. An automatic algorithm detected saccades using minimum velocity and acceleration criteria of 35°s and 9500°s^2, respectively. The radiologists were asked to dictate their findings, which were recorded for further analysis. The viewing time was not limited and varied between 15 to 120 seconds. The included data contains 24 images in which one abnormality was present per image. The images were selected by an experienced radiologist, who also defined the AOI for each abnormality prior to the start of the experiment. The cases were highly diverse and varied in difficulty. The correct diagnosis for each image was known, i.e. confirmed by CT scan or later clinical findings. The verbal responses of the participants were scored by a radiologist and an experienced cognitive psychologist, and were scored as either reported or missed diagnosis. On average, 64.4% of cases were diagnosed correctly. For the illustration purposes, the data analysis using SPLOT focused on comparing the time-courses for looking at the abnormality for the reported and missed diagnosis trials. The analysis was limited to the first 15 seconds of each trial.

3.2 SPLOT analysis

Figure 1 shows the results of the SPLOT analysis applied to the time-courses for the reported and missed diagnosis of chest X-ray. The average time-courses of looking at the abnormality were different, depending on whether the image was diagnosed correctly or not. Participants looked at the abnormality significantly more often when it was reported. Specifically, there were two large significant clusters (both p<.0001), the first one ranging from the start of the trial to about 6 seconds, and the second one from about 7 sec to 15 sec. This shows that the difference between these two conditions emerges very early on and stays at least until the end of the epoch. To analyze whether there were peaks in the looks at abnormality, the time-courses for each condition were compared to their respective average time-courses. This analysis showed 5 clusters for the reported condition, in which only the first two reached significance in a cluster-based permutation testing (p<.0001 and p<.05). This shows a significant peak in the proportion of looks at the abnormality around 1.5 sec from the start of the trial. For the missed condition, there were 2 clusters and only the first one was statistically significant (p<.0001). This shows that there is also a slight increase in looks at the abnormality very early into the trial for the cases in which the abnormality was missed, but after that the proportion of looks at the abnormality remains stable. SPLOT is also compared to *momentous proportion over time,* binned in 60 bins of 250 ms (similar to the kernel size of 100 ms). Same analyses were performed but t-tests with Bonferroni correction were used instead of permutations tests. Although the overall results were similar, many time points did not reach significance. Adding more bins or extending the epoch would render even the strongest effects not significant.

4 DISCUSSION

The present paper describes SPLOT - a new method for analyzing the time-course of looks into an AOI. It provides a complete workflow from visualizing the time-course to performing statistical analysis on it. This straightforward method allows to gain more insight into the time-course of looks at a certain AOI, which is not possible by calculating different summary statistics over the viewing interval (the number of fixations, fixation duration and transitions in and out of AOI). The advantage relative to the most similar *momentous proportion over time* method, is that it preserves the temporal resolution of the eye movement data and uses cluster-based permutation method to perform statistical analysis on the time-course for comparison between conditions and against a baseline. This eliminates the need for correction for multiple comparisons and allows to compare long time-courses. SPLOT is illustrated by the application to eye movements of radiologists and shows that the abnormalities that are reported correctly are looked at very early on (within 1.5 sec). SPLOT has a few limitations. First, it assumes a commonality between looks at the AOI across different images and participants. If the signal is too small or too noisy, no significant differences in looks will emerge. Second, the analysis is focused on one AOI at a time, ignoring transitions between AOIs. When these limitations are kept in mind, SPLOT can provide useful insights in the eye movement dynamics.

ACKNOWLEDGMENTS

This research was supported by an Open Area Research Grant from the Netherlands Organization for Scientific Research to Artem Belopolsky: ORA 464-15-193.

REFERENCES

Richard Andersson, Fernanda Ferreira, and John M. Henderson. 2011. I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica* 137, 2: 208–216.

Floor De Groot, Falk Huettig, and Christian NL Olivers. 2016. When meaning matters: The temporal dynamics of semantic influences on visual attention. *Journal of Experimental Psychology: Human Perception and Performance* 42, 2: 180.

Ralf Engbert and Konstantin Mergenthaler. 2006. Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences* 103, 18: 7192–7197.

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures.* OUP Oxford.

Marcel A. Just and Patricia A. Carpenter. 1980. A Theory of Reading - from Eye Fixations to Comprehension. *Psychological Review* 87, 4: 329–354.

Jonathan van Leeuwen, Jeroen BJ Smeets, and Artem V. Belopolsky. 2019. Forget binning and get SMART: Getting more out of the time-course of response data. *Attention, Perception, & Psychophysics* 81, 8: 2956–2967.

Eric Maris and Robert Oostenveld. 2007. Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods* 164, 1: 177–190.

Ethel Matin. 1974. Saccadic suppression: a review and an analysis. *Psychological bulletin* 81, 12: 899.